A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

PORTUGUESE AUTOMOTIVE AFTERMARKET FORECAST THROUGH TIME-SERIES ANALYSIS AND MODELLING

PEDRO MARIA DE CASTRO LOPES E FIGUEIRA MACHADO | 41020

Work project carried out under the supervision of:

Professor Carlos Santos

04-01-2021

Abstract

The composition of the future car fleet is interesting to many. The automotive aftermarket's concerns rely on the diminishment of profitable segments and the growth of not-so-maintenance-intensive electric vehicles. This project develops time-series forecasting models, namely ARIMA models, predicting the Portuguese car fleet in 2030, which declines 11.6%, to around 5.4 million vehicles. Upon this knowledge, the consequences of the share of electric vehicles rising up to 16.7% are assessed, along with other external factors', under three scenarios. One factor seemingly highly influential is car sharing services, as the vehicle segment typically associated with it is predicted to grow 65%.

Keywords: fleet modelling, automotive aftermarket, forecast, ARIMA

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

A special acknowledgement is in order to my colleague João Pires for his collaboration and companionship.

1. Introduction

Swallowed by a pandemic, businesses struggle to anticipate needs, stay afloat, or, rather simply, survive. One of the sectors that may be deeply influenced is the automotive one, which had already been under recent changes, now faces another variable. What is to expect?

This project consisted in building forecasting models for the Portuguese automotive fleet, as a means of prediction for its composition and, subsequently, its aftermarket value in 2030, while providing more focus on the electric car segment. Historical data was the basis for the forecast, which accounts for entries and exits in the fleet in Auto-Regressive Integrated Moving Average or Brown's Linear Exponential Smoothing models. Results indicate that the fleet should decrease from 6,109,067 vehicles in 2019 to 5,400,598 vehicles in 2030, out of which 901,813 (16.7%) will be electric. This segment is expected to grow a whopping 1900% in the forecasted period. Value-wise, the aftermarket will decrease its value from around 900.5 million euros in 2019 to around 757 million euros in 2030. Another surprising result comes from the growth expected for low cubic capacity diesel vehicles, one of over 65%. This growth may imply that vehicles typically related to car sharing services may be "cannibalizing" other segments, which show significant decline. This type of service is ever more common and will eventually reduce the need for personal vehicles, helping reduce the overall number of cars in the fleet. In the forecast, the impacts of shared car services and legal and environmental frameworks are also taken into account, altering the composition of the fleet by reducing it in 9%, or 486,780 vehicles, or even 142 million euros in aftermarket value, all under a neutral scenario. Additionally, both a pessimistic and an optimistic scenario were assessed. Under these circumstances, the fleet is reduced in 18% and 4.5%, respectively. This would mean doubling (283 million euros) the losses in market value from a pessimistic standpoint, or halving them (71 million euros) from an optimistic one, comparing to a neutral scenario.

Under public pressure and demand for more eco-friendly and sustainable alternatives, the automotive sector has been slowly introducing options that reduce the greenhouse gas emissions or eliminate them altogether. Electric vehicles, henceforth named EVs, such as battery electric vehicles, plug-in hybrid electric vehicles or fuel-cell vehicles (undistinguished in this paper), have been regarded as one of the most promising ones. As estimated by Rietmann, et al. (2020), 30% of the global passenger car fleet will be electric by 2032. As of 2019, EVs occupied only about 1% of the global car stock. Nonetheless, the global electric vehicle fleet has been enlarged significantly in the past decade, mainly due to technology developments and favourable policies (IEA 2020). Besides, price parity between EVs and internal combustion vehicles is on the horizon. According to BloombergNEF (2020), this phenomenon is to have occurred around mid-2020s in most segments, although this may vary widely between geographies. Naturally, price parity is expected to favour EVs competitively, as more eco-friendly means of transportation become financially appealing, too. This steep increase in EVs and their market share is deemed to affect the whole industry. However, one perhaps overlooked component of the automotive sector may have to be especially attentive to this new trend – the aftermarket. Five important changes regarding EVs after-sales services can be identified. These are: decreasing share of mechanical and moving parts, longer service intervals, immature battery technology, less additional units and limited self-service possibility (Dombrowski and Engel 2013). Naturally, the aforementioned alterations to the common vehicle after-sales services that EVs represent are a blow to the automotive aftermarket, which will have to adapt and overcome the losses that may happen, should EVs become increasingly present as is expected. Businesses in the automotive aftermarket can certainly benefit from a better grasp of the future composition of the global, and local, car fleet, allowing them to prepare for future strategies and adaptation, on a level of parts demand, service demand, market value, or business opportunities, as a few examples. Tips4Y is one of the main players in the

Portuguese automotive aftermarket. This automotive intelligence and information systems company provides services to workshops and retailers across the country. The company stands, therefore, on a tightrope regarding the impact of several variables in the automotive aftermarket. Should workshops and parts retailers be affected by the growth of EVs in the fleet, Tips4Y will be affected, too. Naturally, the possibility of acquiring knowledge regarding the future Portuguese car fleets represents an opportunity for the company to adjust strategies, and expect and prepare the unexpected. This paper serves the purpose of shedding light on said knowledge by providing a forecast of the composition of the Portuguese vehicle fleet in 2030, as well as the related maintenance market value. It also hopes to provide related insights, such relations between the prevalence of specific parts in maintenance operations and vehicle segments, and special attention to the electric vehicles segment.

The paper is composed of a literature review on car fleet modelling and the trajectory of electric vehicles, followed by a presentation of the methodology used to execute the proposed tasks. Subsequently, results of said tasks will be analysed and discussed. Finally, further research recommendations will be presented.

2. Literature Review

Due to recent changes and developments in technology, as well as ecologically-driven pressures in a highly-polluting industry, the global car fleet is expected to change in coming years.

In the past couple of decades, concerns with climate change and greenhouse gas emissions have nothing but increased. In fact, as stated by the IPCC in 2007, global carbon-dioxide emissions grew by around 80% between 1970 and 2004 (Braz da Silva and Moura 2016). As accounted in the Portuguese National Inventory Report on Greenhouse Gases 1990-2018 (APA 2020), energy remains the sector with the highest emissions, totalling 72% of total GHG emissions in 2018. Energy industries and transports are the two most important sources inside the energy sector, amounting to 26.6% and 25.6% of total emissions, respectively. This corresponds to around 52.2% of total emissions (72.5% of the energy sector's emissions). IPCC reports have emphasized that should there be no fierce and continuous mitigation policies, transport emissions may increase and reach a faster rate than emissions from the other energy sources in the sector. Therefore, as a means to reduce fossil fuel burning and, thus, slow climate change, several governments have introduced favourable policies to promote the purchase of electric vehicles. (Ellingsen 2016).

As Riemann, et al. (2020) put it, the forecast of sales of electric vehicles is vital for many stakeholders, such as car manufacturers, governments and other policy-making entities, and energy providers, especially for plan-making of electric vehicle production, policy setting and energy grid and supply management. Throughout the literature, it is not uncommon to come across papers whose goal is to predict the composition of the total car fleet. However, works which assess the growth and impact of EVs in the fleet, and therefore, in the environment, are becoming increasingly more common. Ismail and Abu (2013) attempt to forecast demand for a new vehicle to be introduced in a market by means of a Bass diffusion model. This kind of model adopts a rather "marketer" approach, in which adopters of new product are labelled as "innovators" or "imitators", according to how they are influenced, and their interactions dictate the outcome. The basic premise is that the probability of a first purchase on any given time, given that no purchase has been done, is a linear function of the number of previous buyers.

The authors, nonetheless, recognize the limitation of barely relying on historical data, but rather on a set of assumptions. On another hand, Chen (2011) follows a more empirical approach to predict Chinese vehicle demand, building an Auto-Regressive Integrated Moving Average (ARIMA) time-series model which feeds on past monthly sales data. According to the author, ARIMA models are able to attend to a wide variety of situations and provide accurate short-term results. Fridstrøm, et al. (2016) predict the composition of the Norwegian vehicle fleet and its impact in fuel consumption and CO_2 emissions in 2030, by applying a stock-flow cohort model, taking into account two scenarios – applying low carbon fiscal policy, and business-as-usual. By accounting for new registrations, scrapping, and second hand import and export, the authors build the composition of the car fleet. New car registrations follow a discrete choice model, namely a nested logit model. In this particular model, each single car sale is seen as a separate discrete choice, where every car model available in the market is a possibility for the buyer. The nested logit model uses individual vehicle characteristics as explanatory variables. Further on, the information is used to derive intermediate results such as mileage patterns, vehicle survival rates, or life expectancy estimations. Under an environmental policy application, hybrid and battery electric vehicles are estimated to make up for 50% of the youngest vehicle cohort, though reaching only 21% of the total fleet in 2030. The authors also note that the fiscal policy has a large impact in both long term fuel consumption and gas emissions. Laborda and Moral (2020) take yet another interesting and rather different approach. Focusing on forecasting the Spanish automotive aftermarket, they highlight that, since the automobile industry is going through deep transformations, the past, or historical data, is not a solid empirical basis for prediction. Thus, the authors use several regression models to understand which factors related to the automotive industry affect the long term forecasts of the automotive market (using historical data in this step), but then also using participatory methods (using stakeholders' perceptions) to quantify the impact of these conditioning factors, instead of relying on historical data for this second step. Some of the factors taken into account are car-sharing, autonomous cars, electric cars, or legal and environmental frameworks. The authors come to the conclusion that cars with autonomous driving assistance is the factor with most impact, amounting to almost 8,350,000 cars.

On a more specific approach, namely focusing on the trajectory of electric vehicles, Rietmann, et al. (2020) forecast the inventory of this vehicle segment in over 25 countries, while measuring the impact in carbon-dioxide emissions. Using sales data from 2010 to 2018 and a logistic growth model, the authors forecast on a 15-year horizon, mentioning the 30% of the total fleet threshold that EVs will occupy in 2032. Country-wise, the paper shows how different countries will see different results depending on certain current conditions. Countries such as Norway, Switzerland or Belgium should see fast EV adoption and high CO_2 emissions reduction mainly due to their government policies and goals, as well as cleaner energy mixes. On the other end of the spectrum, China and India are predicted to see slow EV penetration and even a strong CO_2 emissions increase. This result is attributed to low government support in adopting EVs but also to the fact that both countries still have a long way to go in what clean energy generation is concerned. In summarized results, although EVs will be 42.5% of the total car fleet across the 26 countries in study, carbon-dioxide emissions will still rise by 11.8% between 2018 and 2035. Nonetheless, the authors emphasize that EV adoption is still a good solution, as an entirely internal combustion engine car fleet would see CO_2 emissions rise 29% until 2035. Braz da Silva and Moura (2016) apply a systems dynamics model to the Portuguese car fleet in order to capture the relationships between the drivers of the system, while simulating drivers' purchase behaviours. Some of the drivers are: demography and GDP, motorization rate, energy and vehicle costs, taxes and subsidies, composition of the fleet (both vehicle and fuel type) and annual mileage of vehicles. A combined aggregate time series with a static disaggregate discrete (car-types) choice model is used to assess the changes in the fleet. Results show that the diffusion of battery electric vehicles in the market will not exceed 7.51% of the total Portuguese vehicle fleet by 2030, while hybrid vehicles could reach up to 59,81%.

3. Methodology

Initially, relevant data was collected that could be used for the purposes of forecasting the automotive aftermarket by means of forecasting the composition of the fleet in 2030.

Concerning data collection, Tips4Y granted access to two sets of data. The first one consisted of a dataset of queries made to Tips4Y's Vehicle Running Costs (VRC) platform. This platform allows users to check an individual vehicle's maintenance scheduled by manufacturers and its respective budget by introducing the vehicle's license plate and its mileage. Each query generates an entry in Tips4Y's database, where the queried vehicle's characteristics (model, power, cubic capacity, gross weight, etc.) are also included. Thus, a dataset containing information on 5,739,177 unique vehicles was obtained. It is important to note that electric vehicles were removed from this dataset and were included in the analysis later on from a different source. The second set of data consisted of 990 sample vehicle entries, again, containing several of the vehicles' characteristics. Since this dataset was representative of the whole fleet, it would be used to segment vehicles according to their budgeted maintenance cost. The focus of the project was the automotive aftermarket, thus, it made sense to act according to its perspective. None of the datasets included the vehicles' respective maintenance costs for the scheduled operation. Therefore, these values were manually retrieved for the dataset of 990 vehicles from Tips4Y's VRC platform. Mileage was missing information and had to be inferred in order to use the platform. In order to do so, a study by Observatório ACP (2018) in which a relationship between a vehicle's age and mileage was described (Appendix 1). Therefore, the maintenance costs for the next scheduled maintenance operation were retrieved for each of the 990 sample license plates. After removal of invalid values, 716 entries remained. Subsequently, after obtaining maintenance costs for 716 vehicles, the segmentation step followed. It was, therefore, important to analyse which characteristics were prevalent in influencing and determining maintenance costs. The characteristics in study were: Cubic Capacity, Gross Weight, Gearbox (e.g. manual, automatic), Vehicle Type (e.g. passenger vehicle, commercial vehicle), Category (e.g. light-duty vehicle, heavy-duty vehicle), Body Style (e.g. SUV, sedan, coupé) and Fuel Type (e.g. diesel, petrol, hybrid). The analysis was conducted through a decision tree, generated using the statistical analysis tool SPSS Statistics, which can be defined as a procedure that recursively partitions data into smaller subdivisions with splits determined by a statistical procedure (Friedl and Brodley 1997). In this particular case, the statistical procedure was the CHAID (Chi-square Automatic Interaction Detector) growth method, in which data is split according to statistically significant chi-square tests for independence on categories of predictor variables. This method allows for splitting data into more than two categories (multi-way split). This characteristic provided a more sensible and interpretable approach while not extending tree depth, instead of a two-way split only model, and was therefore chosen for this reason over other alternatives. The generation of several trees was conducted, changing criteria such as case limits or tree depth limits, with further comparison of the results. The tree with most conceptual sense and interpretability was chosen. From this procedure, eight segments were created, with basis on Fuel Type, Cubic Capacity and Vehicle Type. A ninth segment is included later, consisting solely of electric vehicles. Once the characteristics of each segment were assessed, the segmentation was applied to the 5,739,177 vehicles in the main dataset. Once this application was done, the data was grouped by segment and license plate year, corresponding to the year the vehicle entered the fleet. The result was a table with the number of entries in each year (from 1990 to 2020) by segment. Data for 2020 was discarded as it represented only the first few months of the year.

Moving on to the second step of the model, the goal was to forecast the number of vehicles entering the fleet in all years from 2020 (inclusively) to 2030. For this purpose, an Auto-Regressive Integrated Moving Average (ARIMA) model was used. The ARIMA has been a popular linear model for time-series forecasting for the last few decades. In this model, the future value of the variable in study is a linear function of past values (or lags) and random errors (Zhang 2003). The model can be translated by the following general equation:

$$\hat{y}_t = \mu + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \tag{1}$$

where $\hat{y}$ is the estimated value at time *t*, ε_t is the random error at time *t* and μ is the model's constant. φ_i and θ_i are model parameters, while *p* and *q* correspond to the model's orders. *p* represents the number of autoregressive terms in the model and *q* accounts for the number of lagged forecast errors (representative of the moving average component). The ARIMA model includes yet another order *d* related to the number of non-seasonal differences that are needed to reach a stationary series. Hence, an ARIMA(*p*, *d*, *q*) model requires a stationary time-series where random errors are independently and identically distributed with a mean of zero and constant variance. The series, therefore, had to be, and were, tested for stationarity using Dickey-Fuller and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests for stationarity. Should they have proved to be non-stationary, differencing had to be applied, stabilizing the mean over one constant value, which removes trend (e.g. an ascending pattern) in the series. Another kind of operation used to provide stationarity was applying the natural logarithm of the series, which tends to stabilize the variance (e.g. different amplitudes of oscillating around a mean value) around the mean of the series. Once stationarity was achieved, analysis of the autocorrelations and partial autocorrelations of the series provided insights on which parameters to use and what models were adequate. This procedure yielded statistically significant models with reasonable forecasts. For the ninth segment, comprising of electric vehicles, an additional model was developed, in order to consider two scenarios. The model developed for a more conservative scenario of entries of these vehicles in the market was a Brown's Linear Exponential Smoothing model. Unlike a simple exponential smoothing model, this model accounts for trend in the data and applies an exponential filter twice, generating a linear forecast and not an exponential one.

On a third step, the number of market exits was assessed. Naturally, vehicles have a survival rate, meaning that on any given year, they might exit the market due to several reasons, such as scrappage, deregistrations, or accidents. Survival rate is defined as the probability of a vehicle staying in the fleet. Based on the work of Fridstrøm, et al. (2016), the average survival rates were computed for each year of age of a vehicle (Appendix 2). This rate, however, is based solely on scrappage and deregistrations. It is important to note that for ages over 31 years a constant value equalling that of 31 years of age was assumed. Segment market exits at any time after *t* years since the beginning of the analysis follow the following series:

$$\text{Market Exits } (t) = \sum_{i=1}^{t} \text{Market Entry}_{i-1} \text{ (1-Survival Rate}_{t-i+1}) \tag{2}$$

One must note, however, that Market Exits in year 0 are also 0. It is assumed that during the first year of its lifetime, no car exits the market. With both market entries and market exits for all years in study, the number of vehicles of a segment at every year was calculated by simply adding these two sets of values. After having the evolution of each segments' fleet over the years, the segments were compounded in order to reach the total fleet's evolution. Having the composition of the fleet in 2030, the annual business volume of the automotive aftermarket in Portugal was calculated with basis on the mean maintenance costs for each segment, which were part of the output of the decision tree where segmentation was characterized. One segment was missing in this analysis, which is the segment of electric vehicles, thus, its average maintenance cost had to be sourced elsewhere. According to Propfe, et al. (2012), plug-in hybrid electric vehicles spend 26.5% less in maintenance than internal combustion vehicles. Therefore, taking into account our estimates of the average maintenance costs for these vehicles, the average maintenance cost for the electric vehicle segment was inferred. According to the decision tree algorithm, internal combustion vehicles (including hybrid vehicles in very low percentage) in the dataset in study have an average maintenance cost of 291.52 euros. Therefore, having 26.5% less costs, electric vehicles were assumed to have an average 214.27 euros in maintenance costs.

4. Analysis and Results

The first step of analysis consisted of generating a decision tree that provided reliable characterization of the automotive segments, grounded on their average maintenance costs. As seen in the figure below, the algorithm split the data due to differences in the variables Fuel Type, Cubic Capacity and Vehicle Type.

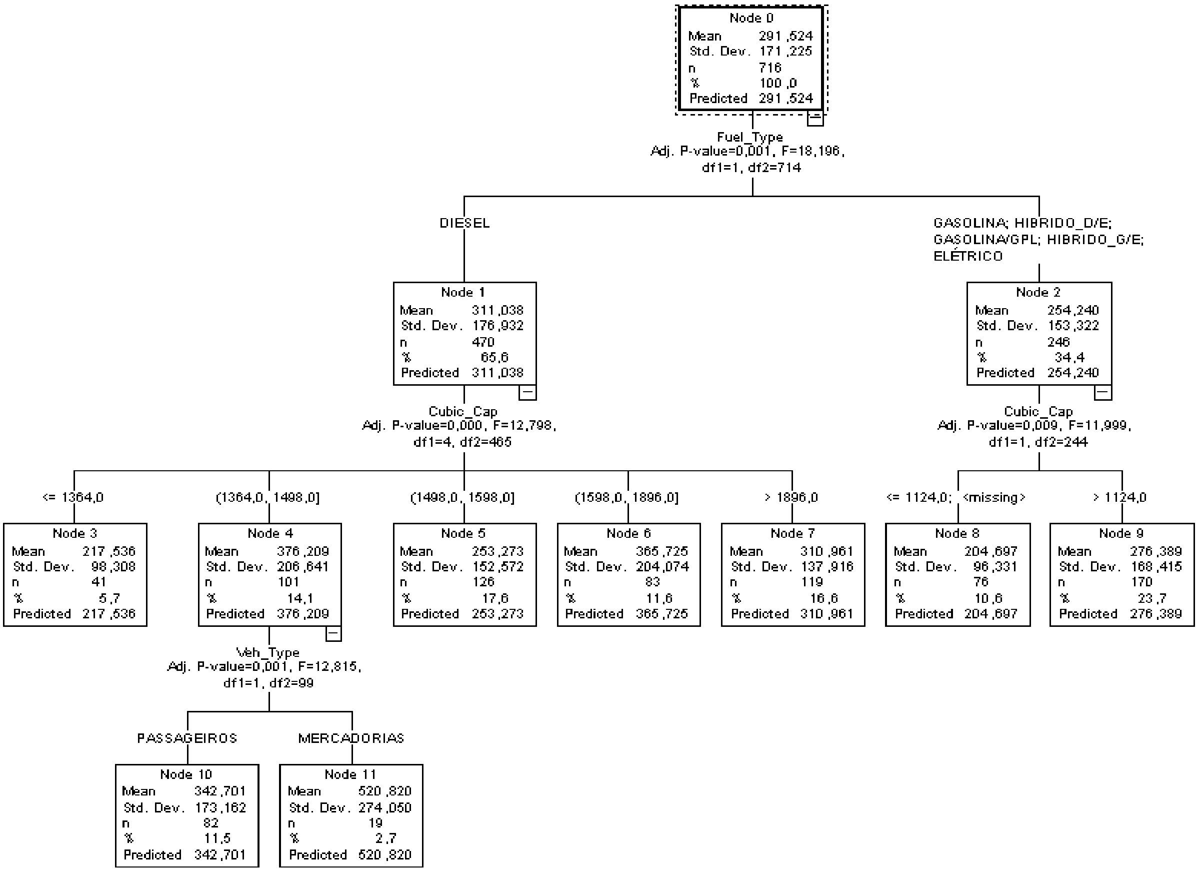


Figure 1 - Decision Tree characterizing segments[1]

The algorithm found Fuel Type to be the most relevant variable to characterize maintenance costs. Interestingly, diesel vehicles were found to have higher maintenance costs than all other types of fuel. The second most important variable was Cubic Capacity. It is expectable that vehicles with higher cubic capacity and power have higher costs. However, evidence suggests

[1] The decision tree attempts at segmenting vehicles according to their maintenance cost differences. Node 0 contains all sample vehicles, which are progressively split into nodes according to Fuel Type, Cubic Capacity and Vehicle Type. On each node, a mean maintenance cost is displayed. Each terminal node represents a segment.

that vehicles with Cubic Capacity between 1364 and 1498 cubic centimetres have higher maintenance costs than all others. This may be due to sampling issues, in which vehicles of said Cubic Capacity were predominantly of a higher-end manufacturer and, thus, parts and maintenance operations could be more expensive, driving maintenance costs up. The same thought process could be considered for vehicles with Cubic Capacity between 1598 and 1896 cubic centimetres. Apart from these two groups, higher Cubic Capacity revealed higher maintenance costs. Concerning vehicles with Cubic Capacity between 1364 and 1498, a further split was done, in respect to the Vehicle Type. Naturally, commercial vehicles, which tend to be used more often and, thus, go more often to the workshop or replace parts more often are expected to have higher costs due to the higher wear of the vehicle. Based on the evidence suggested by the decision tree output, eight segments were created, each corresponding to vehicles with characteristics described by each of the tree's final, split-less, nodes (leaf nodes). An additional segment of electric vehicles was created, as these vehicles were under represented in the original dataset, having only one case (about 0.14%). Segments are summarized below.

Table 1 - Summary of segment characteristics[2]

Segment	Decision Tree Node	Fuel Type	Cubic Capacity	Vehicle Type
Segment 1	3	Diesel	≤1364	-
Segment 2	10	Diesel	]1364, 1498]	Passenger
Segment 3	11	Diesel	]1364, 1498]	Commercial
Segment 4	5	Diesel	]1498, 1598]	-
Segment 5	6	Diesel	]1598, 1896]	-
Segment 6	7	Diesel	>1896	-
Segment 7	8	Petrol/Hybrid	≤1124	-
Segment 8	9	Petrol/Hybrid	>1124	-
Segment 9	-	Electric	-	-

[2] Vehicle segments are defined by the combination of their fuel type, cubic capacity and vehicle type. Each segment is represented in a decision tree terminal node. A ninth segment was added, comprising of only electric vehicles, which were underrepresented in the sample.

Having segmented the main dataset and compiling data, the forecasting model development was initiated. This process was done through use of an ARIMA(*p*,*d*,*q*) model. Since ARIMA models require stationary time-series, each of the segment series was tested for stationarity through Dickey-Fuller and KPSS tests. Results indicated that some segments, such as segments 3 through 7, had stationary series and, therefore, required no differencing. On the other hand, some other segments did require more attention on this matter. Thus, several differencing terms were tested on all models, in order to find the most accurate model, both in terms of fit and reasonability of predictions. The models chosen and their results and statistics are summarized in the table below.

Table 2 - Characterization of ARIMA models used[3]

Segment	Model	Natural log	Constant	R^2	Ljung-Box test	BIC	AR p-value	MA p-value
1	(2,2,1)	No	No	0.836	0.845	17.064	Both 0.000	0.005
2	(0,1,1)	Yes	No	0.728	0.768	18.068	-	0.022
3	(1,0,1)	No	No	0.873	0.591	15.090	0.000	0.000
4	(1,0,0)	No	No	0.840	0.931	18.062	0.000	-
5	(1,0,1)	No	No	0.932	0.147	17.535	0.000	0.028
6	(1,0,0)	No	No	0.901	0.063	18.001	0.000	-
7	(1,0,1)	No	Yes	0.875	0.668	16.860	0.000	0.056
8	(1,1,0)	No	No	0.937	0.269	18.150	0.003	-
9	(0,2,0)	No	Yes	0.974	-	13.448	-	-

Analysing the results, one can see that only the models for segment 1 and 9 required second-order differencing, in which differencing is applied to the already differenced series. Most models had one term of auto-regression, while a moving average term was about as common as having none. Models for segments 2 and 9 had no auto-regression factor, meaning that future

[3] ARIMA(p,d,q) models used for forecasting are defined by their three components, logarithmic transformations to the data, and the presence of a constant. The significance and performance of the models is evaluated according to statistics such as the r-squared, Ljung-Box test statistic, Bayesian Information Criterion and individual parameter significance values.

values do not depend on previous lags. Overall, the results provided reasonable estimates, with adequate fits and statistically significant parameters. It is, however, pertinent to note that the model for segment 6 had a particularly low p-value for the Ljung-Box test. This test assesses whether autocorrelations in the series are statistically different from zero and errors are a product of randomness. The null-hypothesis of the test states that the data is independently distributed. Therefore, a positive result would be to fail to reject the null-hypothesis, or rather, yielding a p-value above significance level. For a confidence level of 95% (hence, a significance level of 0.05), the model for segment 6 barely made the cut, which could reveal hindering autocorrelations. Nonetheless, the model had positive results and was kept. On another note, the MA parameter of the model for segment 7 failed to reveal itself statistically different from 0. The p-value was right about the established 0.05 significance level, though. Since results were also positive, the model was kept. Finally, a normalized Bayesian Information Criterion (BIC) was analysed in order to compare models for the same time-series. This criterion evaluates overfitting due to excess of parameters. The lower the value, the better the model would be, comparatively to another with higher normalized BIC value. The forecasting yielded from the models above provided reasonable results, overall, with adequate values for the dependent variable. Therefore, market entry values were predicted and set according to them for the forecast period. However, in order to assess results from another procedure, as a way to rule out the possibility of having inadequate results, as well as provide another possibility of choice, a different kind of model was developed for segment 9. This segment was lacking in data compared to the others and could have been providing unreasonable results, from an interpretative standpoint, being possibly too optimistic. Thus, a more conservative approach was taken as a means to provide an alternative to choose from, should it prove to be more reasonable. Because the ARIMA model for segment 9 generated an exponential curve in its forecast, it was decided to test an exponential smoothing model, in particular, Brown's linear exponential smoothing model. The model included no transformations to the data and returned an alpha parameter with p-value equal to 0.000, along with a value for R-squared of 0.959. As expected, the model provided a more conservative forecast, as the curve is, in fact, linear and not exponential. However, the ARIMA model still yielded positive, reasonable and comparatively better results. Therefore, for the continuation of the analysis, in particular, the composition of the fleet, the forecast computed by the ARIMA model was taken into account. Plots of both models can be found in Appendix 3 and 4.

The evolution of both total entries and total exits in the fleet is depicted in Figure 2 below.

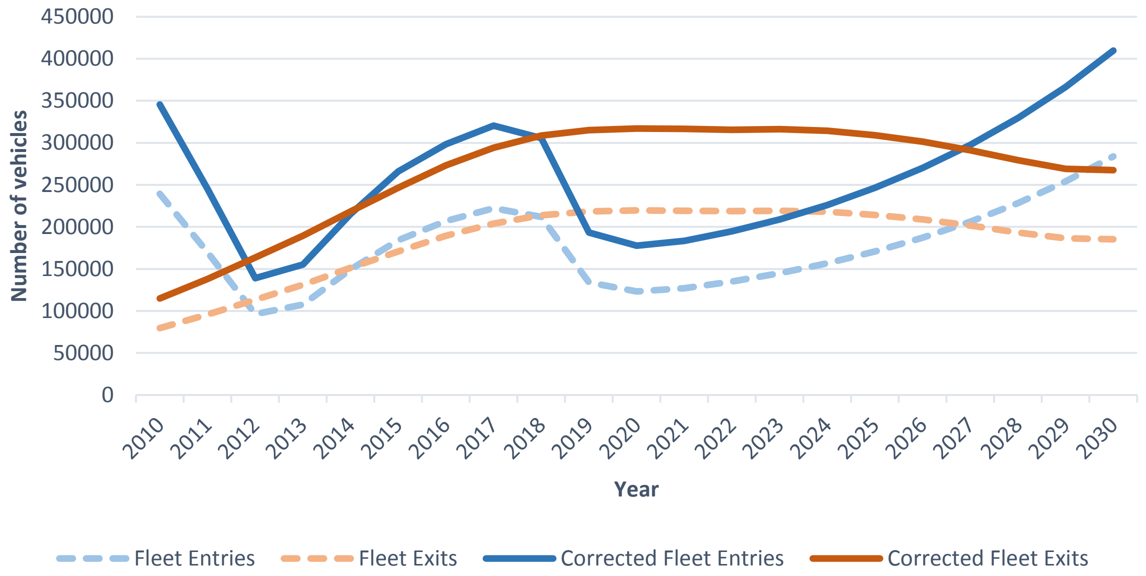


Figure 2 - Total fleet entries and exits across 20 years, including forecast[4]

As one can see, fleet entries have followed a more erratic pattern, dropping significantly in 2012-2013, climbing in recent years, only to fall again in 2019. Due to the economic crisis felt in 2012 in Portugal, it is reasonable to accept such a decline in market entries in said period. However, the sudden drop of entries in 2019 proves to be more intriguing. The reason for this drop may lie in the fact that the data retrieved and analysed is from early/mid 2020. It is

[4] The figure displays values for both fleet entries and exits which were corrected due to lack of representability by applying an overall proportionality factor. Values from 2020 to 2030 are the product of a forecast.

important to recall the fact that the presence of vehicles in the dataset is dependent on a vehicle having been queried in Tips4Y's VRC platform. Naturally, cars entering the market in 2019 probably have not yet been to a workshop for maintenance or repair, nor have been queried in the VRC platform, as light-duty passenger vehicles usually have maintenance operations scheduled every two years. Therefore, it is reasonable to suppose that the dataset in study may be severely lacking in representation of 2019 vehicles, hence suggesting low entries in that year. The forecast, however, remains optimistic, with entries rising to values higher than the past 10 years in 2030.

Because our vehicle data is dependent on VRC platform's queries, it was decided to investigate the matter and understand the representability of our data in relation to official data. Therefore, fleet entry data from PORDATA (2020), one of Portugal's main statistical databases, was collected and compared. It was found that, on average, our entry data represented only 69% of the total fleet entries in Portugal, with the values for 2019 representing a hindering 38%. Concerning this issue of low representability, the fleet entry values for each year were adjusted according to the average 69% proportionality factor. The same approach was used for the total fleet values, where the data represented 63% of official values, though variation around this number was relatively small (2019, the lowest, represented 54% of the corresponding year's actual total fleet). With all values adjusted, it was possible to proceed the analysis. It is important to note that this correction adjusts the overall scale of the analysis, but not specific year-wise patterns, such as the drop in 2019. This would require year-by-year corrections which would render the initial data irrelevant, and therefore were not applied. The evolution of the total fleet can be visualized below.

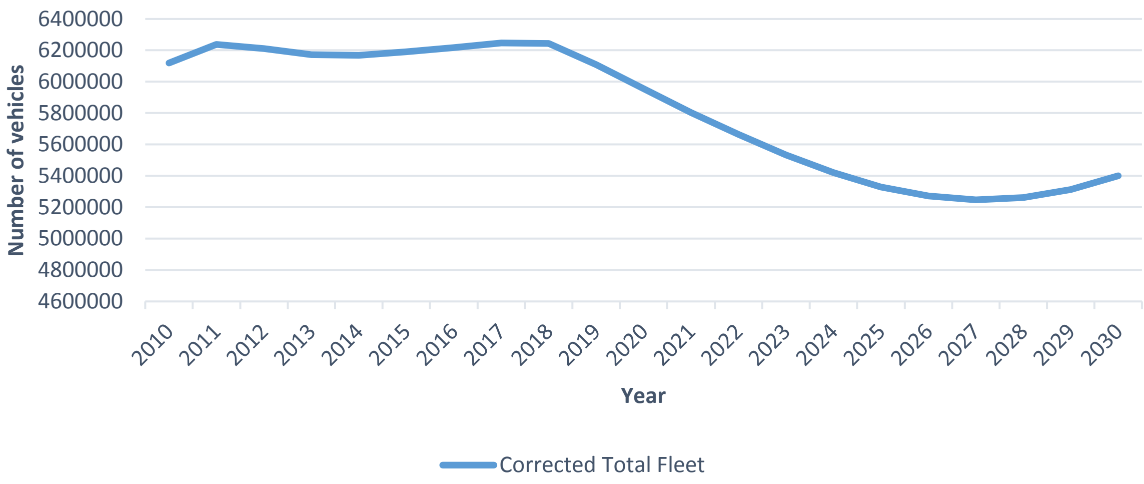


Figure 3 - Corrected Evolution of total Portuguese fleet, including forecast[5]

According to our estimates, the situation looks grim. After a relative plateau in the number of vehicles circulating in Portugal from 2010 to 2019, fleet exits start to exceed fleet entries, resulting in the descending curvature depicted. However, around 2026-2027, the situation is inverted again. A total of 5,400,598 vehicles is forecasted for the Portuguese car fleet in 2030. It is possible that the sudden drop in entries in 2019 is to blame for a pessimistic forecast in the upcoming years, as the models generate values based mostly on the previous lag, therefore, a low value in the last observed year may tend to provoke a descending slope in forecasting. As the proportionality factor applied was the same (the average) throughout the years, the fall in 2019 remained.

[5] The figure displays the evolution of the total Portuguese car fleet from 2010 to 2030. The later 10 years are the product of a forecast, while the earlier are based on historical data. All values have been corrected with a proportionality factor due to lack of representability of real, well-documented values

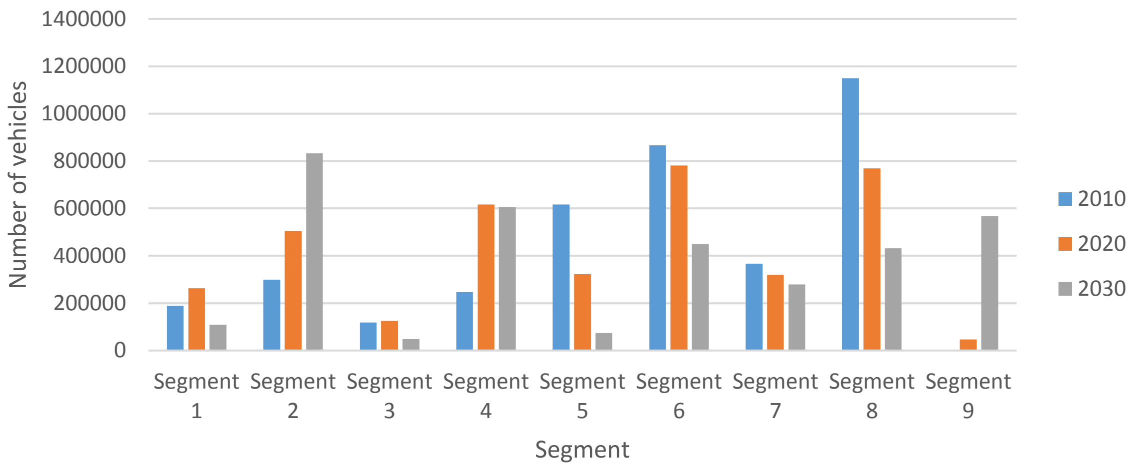


Figure 4 - Evolution of the number of vehicles per segment, in 10-year periods[6]

The figure above shows how only two out of the nine segments are predicted to increase in number. Except for segment 4, all shrinking segments are reduced to numbers lower to that of 2010. Segments 1 to 6, which consist of diesel-powered vehicles with ordered different cubic capacities, seem to indicate a decline, generally, in the use of this type of highly polluting vehicles. The exception seems to be segment 2, consisting of rather low cubic capacity diesel vehicles, which is expected to increase in coming years. This may be an indication of an increase in car sharing services, which may tend to reduce other segments (as fewer cars will be needed) but average medium-quality vehicles may increase in demand, especially from this kind of activity. Petrol vehicles, represented in segments 7 and 8, with low and high cubic capacities respectively, seem to be in decline as well. While petrol is generally more expensive than diesel, this kind of vehicles is, too, very polluting. These results may imply a progressive change from internal combustion engines to electric and hybrid ones. The latter are underrepresented but may, in fact, become a more present alternative in the future. Although these vehicles apparently share the same maintenance costs as petrol ones, they may be

[6] The figure depicts both the historical and forecasted evolution of vehicle segments in fleet size, over the span of two decades under 10-year periods. Segments 2 and 9 display strong growth, while others decline.

significantly more eco-friendly and technologically enhanced in the near future, thus shrinking the numbers of pure internal combustion engines. Another interesting note will be to see how the electric vehicle segment, segment 9, will play out in the future. According to our estimates, this will be the third most represented segment in the Portuguese fleet, accounting for 16.7% of the total fleet's vehicles. Concerns towards climate change, public pressure, and appreciation of these modern vehicles have been tending to drive sales of electric vehicles up in an accelerating fashion. Our model represents this trend, forecasting the presence of 567,562 electric vehicles in Portugal in 2030, with 105,392 entries in the same year. These numbers don't appear unreasonable and may reveal the effect of electric-friendly policies in place in Portugal.

Lastly, with the purpose of evaluating the impact of some of the factors that condition the composition of the car fleet of the future, the work of Laborda and Moral (2020) was consulted. The authors estimate the impact of several factors in the number of vehicles of the 2030 car fleet in Spain. Under three scenarios – pessimistic, neutral, and optimistic – the increasing presence of car sharing services, and legal and environmental frameworks in place were the matters in study. Under the neutral scenario, Laborda and Moral (2020) expect a reduction in around 2.39% and around 6.62% of the total fleet in 2030, due to car sharing and legal and environmental frameworks, respectively. In Portugal, this would amount to 129,096 and 357,684 less vehicles in circulation. A combined reduction of 486,780 vehicles (around 9%) would result in 4,913,819 vehicles. This is a very significant impact and reflects the threat that these factors may pose to the automotive aftermarket. Figure 5 below sheds light on other scenarios' impact, as well as how they relate to the initial forecast.

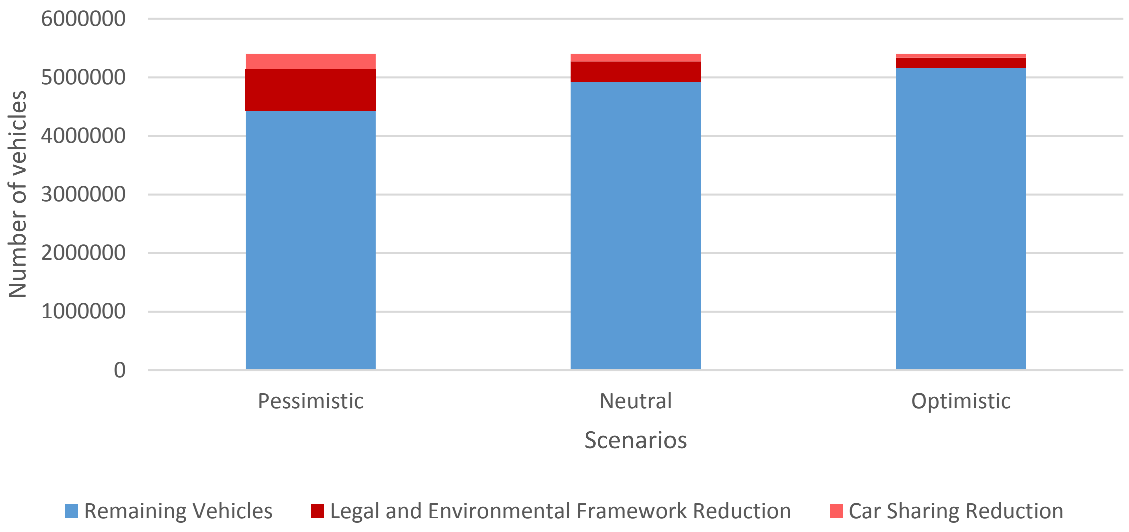


Figure 5 - Car Sharing and Legal and Environmental Frameworks impact on Total Fleet[7]

One of the topics of interest for the automotive aftermarket sector is the prevalence of specific parts in specific segments. This study also focused on what parts were more likely to be required for each segment. When it comes to worn out parts that might or might not be replaced in maintenance operation, windshield wipers are by far the most common across all segments, reaching an average 94.7% of all vehicles. Segment-wise, segment 4 (consisting of diesel-powered medium cubic capacity vehicles) is the one with fewest replacements on worn out parts, meaning it may be considered the most "durable" segment. Concerning parts that are mandatorily replaced on a scheduled maintenance operation, regardless of its apparent state, engine oil, engine oil filter, and cabin filter are the most commonly replaced ones, reaching 94.4%, 93.9% and 66.5% of all scheduled operations, respectively. From the automotive aftermarket perspective, specifically in parts replacement, segment 3 (commercial vehicles

[7] The figure displays the impact of applied conditioning factors in total fleet size. Both car sharing services and legal and environmental frameworks contribute to a decrease in the number of vehicles by 2030. Three scenarios are displayed, according to the severity of the factors.

powered by diesel with relatively low cubic capacity) proves to be the most profitable, with most scheduled replacements.

5. Discussion

From the results that were obtained, one can see an expected decline of almost all segments, which may indicate either a progressive switch to other more ecological means of transportations, such as public transportation, bicycles, or others, or a continuous change towards segments which include eco-friendlier alternatives, which seem to growing. Segment 9, comprising of electric vehicles only, is expected to grow exponential in coming years, which only goes according to the trend of recent years, as it has grown over 7500% in registers from 2010 to 2019. Segment 1, diesel vehicles with lowest cubic capacity, are expected to decrease numbers continuously. What is most interesting, however, is how these cars seem to stop being registered, or eventually sold, from 2021 onwards. Usually, diesel vehicles are associated to higher cubic capacity and more potent engines. If so, naturally, it would be adequate to assume increasing the redundancy of these vehicles which have lower cubic capacity. Segment 2 (similar vehicles but with slightly higher cubic capacity), however, is expected to grow. A possible explanation might be the common presence of these vehicles in car sharing, which requires vehicles capable of enduring long distances without much stop. Diesel engines are usually sturdier than the petrol ones, with lower fuel costs, as well as allowing longer autonomy compared to electric vehicles. According to our maintenance costs estimates, this segment has significantly higher costs than almost all of its diesel counterparts, indicating higher usage. Cost-wise, the exception is segment 3 (vehicles equal to those of segment 3, only of the commercial type). This segment naturally has higher costs due to more usage, as commercial vehicles tend to be used more often and for longer distances and durations. This segment is expected to decline, as well. Perhaps the additional costs undermine its advantages compared to segment 2's vehicles, and users might tend to switch towards passenger type vehicles, even for other commercial purposes. While segment 4, of medium cubic capacity diesel vehicles will decline very little, segments 5 and 6 (diesel vehicles of highest cubic capacity) will decline very significantly. Higher cubic capacity and potency usually relates to higher emissions, turning powerful vehicles increasingly unattractive to newer and more climate-concerned generations. Segment 4, nonetheless, may still possess the characteristics to remain mildly attractive, having enough cubic capacity to make sense for a diesel vehicle, but not enough to cause too much environmental distress in users. Besides, a lower fuel price compared to petrol vehicles may be an incentive towards diesel vehicles for many people if ever comparing diesel and petrol vehicles. Petrol-wise, all vehicles are expected to decline. Higher fuel prices probably undermine the lower maintenance costs. Higher cubic capacity vehicles seem to decline more abruptly than lower ones, in the coming years. Again, environmental reasons associated with higher potency may be the reason.

With the composition of the Portuguese fleet in 2030 forecasted, it is pertinent to calculate the amount of business that the automotive aftermarket may expect from a maintenance standpoint. Basing the approach on the output values for the mean maintenance costs provided by the decision tree employed earlier, it was possible to reach a value of around 757 million euros in 2030, compared with 900.6 million euros in 2019. Notably, the electric vehicle segment multiplies its value by almost 20-fold, from a bit less than 5 million euros to about 96.6 million euros. Besides segment 9, also segments 2 and 4 are expected to generate more value in the future. Segment 2 may bring about almost another 90 million euros in value, while segment 4 might increase volume in a more modest 2 million euros. The growth of these sectors cannot offset the decrease of the others. This may be demotivating news for the automotive aftermarket that may have to adapt to accommodate the new-coming electric vehicles and make use of the business that they may bring, or suffer severe losses in the growing share of electric vehicles. On yet another negative remark, conditioning factors (car sharing and legal and environmental frameworks) are also deemed to reduce value in the market. On a neutral scenario, these factors are expected to decrease the value of the market by about 142 million euros, which can be aggravated up to 283 million euros under the pessimistic scenario. However, though car sharing may reduce the numbers in the fleet, vehicles in such activity may increase maintenance shops' revenues by 40%, according to experts' opinions (Laborda and Moral 2020).

6. Limitations and Recommendations

As with any piece of research, there are limitations to be exposed. Although the ARIMA models provided decent forecasts, these are very simple time-series forecasting models that may fail to recognize causalities in the data, such as relations between segments, seasonality, and other intricacies. More complex models may prove to be more successful in this task. Moreover, the maintenance costs for each of the 716 license plates available were retrieved online manually. Naturally, this process may be significantly influenced by human error, although analysis of outliers is helpful in detecting this. Besides, relevant data concerning electric vehicles was lacking, thus their average maintenance cost had to be inferred based on assumptions, instead of empirically as was done with the other segments. The world faces constant change and historical data may not always be the most reliable piece of information. Currently, we are facing a pandemic that is sure to cause impact in the automotive sector. Further studies would benefit from accounting for this external factor, along with others such as autonomous vehicles and driving-assisted vehicles. Another hindering condition has been the source of the data that was used for this project. As has been mentioned previously, the main dataset in study is composed of a platform's queries registers, in which corresponding vehicle data is associated. Naturally, this limits the sample to vehicles that have been once queried, either this year or 10 years ago. Some of those vehicles may not even be in circulation as of this year. Besides, recent vehicles are bound to not be present in the dataset as having been to a workshop, or rather, been queried in the platform is a "requirement" for it. A lack of vehicles in the most recent observed case proves to highly influential of the outcome of forecasts, as these rely mostly on the previous records, especially those closer, in chronological terms, to the predicted value. Therefore, further studies with similar goals may make good use of a dataset consisting of the totality of currently registered vehicles, in order to avoid misleading conclusions.

7. Conclusion

Throughout this project, the main goal was to assess the threats that the future may pose for the automotive aftermarket in Portugal. By developing and testing several forecasting models for the composition of the car fleet in 2030, results indicated that Portugal will have 16.7% of its car fleet consisting of electric vehicles. Braz da Siva and Moura (2016), however, predict 7.51% of electric vehicles in the same country. Opposite, Rietmann et al. (2020) predict a market share of EVs of 36.2% in Portugal in 2030. This paper's results, apparently, stand in-between these authors. One might think that the more optimistic results may come from the fact that the analysis was conducted later, with more information on the exponential growth of EV sales. Nonetheless, and despite the uncertainty, it was possible to retain insights concerning the business volume that the market may have in the future, aiding in strategy definition and adaptation policies. Electric vehicles were predicted to occupy a significant portion of the future fleet, along with implications that this fact may bring. Adding to other automotive-related pieces of information, the impact of a couple of external factors was studied. Finally, taking into account possible limitations of the project, several recommendations are put in place, to further contribute to research in this topic and evidence-based decision-making in the automotive sector.

References

Rietmann, Nele, Hügler, Beatrice, Lieven, Theo. 2020. "Forecasting the trajectory of electric vehicle sales and the consequences for worldwide CO_2 emissions". *Journal of Cleaner Production,* 261

International Energy Agency. 2020. "Global EV Outlook 2020", Accessed December 2020. https://www.iea.org/reports/global-ev-outlook-2020

Bloomberg New Energy Finance. 2020. "Electric Vehicle Outlook 2020", Accessed December 2020. https://about.bnef.com/electric-vehicle-outlook/

Dombrowski, Uwe, Engel, Christian. 2013. "After Sales Strategies for the Original Equipment Manufacturer of Electric Mobiles". In *Re-engineering Manufacturing for Sustainability,* edited by Andrew Y. C. Nee, Bin Song, and Soh-Khim Ong, 347-352. Singapore: Springer.

Tips4y. 2020. Factos e Números, Accessed December 2020. https://www.tips4y.pt/

Ellingsen, Linda Ager-Wick, et al. 2016. "The size and range effect: lifecycle greenhouse gas emissions of electric vehicles". *Environmental Research Letters,* 11

Braz da Silva, Marta, Moura, Filipe. 2016. "Electric vehicle diffusion in the Portuguese automobile market". *International Journal of Sustainable Transportation,* 10(2): 49-64

APA (Portuguese Environment Agency). 2020. "Portuguese National Inventory Report on Greenhouse Gases, 1990-2018". Accessed December 2020. https://unfccc.int/documents/215705

Ismail, Zuhaimy, Abu, Noratikah. 2013. "New Car Demand Modeling and Forecasting Using Bass Diffusion Model". *American Journal of Applied Sciences,* 10(6): 536-541

Chen, Daoping. 2011. "Chinese Automobile Demand Prediction Based On ARIMA Model". *4th International Conference on Biomedical Engineering and Informatics (BMEI)*

Fridstrøm, Lasse, Østli, Vegard, Johansen, Kjell Werner. 2016. "A stock-flow cohort model of the national car fleet". *European Transport Research Review,* 8

Laborda, Juan, Moral, Maria José. 2020. "Automotive Aftermarket Forecast in a Changing World: The Stakeholders' Perceptions Boost!". *Sustainability*

Observatório Automóvel Clube de Portugal. 2018. "Estudo Condutor Português". Accessed November 2020. https://observatorio.acp.pt/estudos/condutor_portugues/12/index.html

Friedl, M.A., Brodley, C.E. 1997. "Decision tree classification of land cover from remotely sensed data". In *Remote Sensing of Environment,* 61(3), 399-409. USA: Elsevier

Zhang, G. Peter. 2003. “Time series forecasting using a hybrid ARIMA and neural network model”. In *Neurocomputing*, 50, 159-175. USA: Elsevier

Tabachnick, Barbara G., Fidell, Linda S. 2013. “Time-Series Analysis”. In *Using Multivariate Statistics, 6th Edition*. USA: Pearson

Propfe, Bernd, Redelbach, Martin, Santini, Danilo J., Friedrich, Horst. 2012. “Cost analysis of Plug-in Hybrid Electric Vehicles including Maintenance & Repair Costs and Resale Values”. *World Electric Vehicle Journal*, 5: 886:895

PORDATA. 2020. “Veículos rodoviários motorizados em circulação: total e por tipo de veículos”. Accessed December 2020. https://www.pordata.pt/Portugal/Ve%c3%adculos+rodovi%c3%a1rios+motorizados+em+circula%c3%a7%c3%a3o+total+e+por+tipo+de+ve%c3%adculos-3100

PORDATA. 2020. “Veículos matriculados: total e por tipo de veículo”. Accessed December 2020. https://www.pordata.pt/Portugal/Ve%c3%adculos+matriculados+total+e+por+tipo+de+ve%c3%adculo-3103

Monteiro, João. 2019. “Forecasting the Portuguese Car Fleet and Automotive Aftermarket for the future”. Accessed October 2020. http://hdl.handle.net/10362/104575

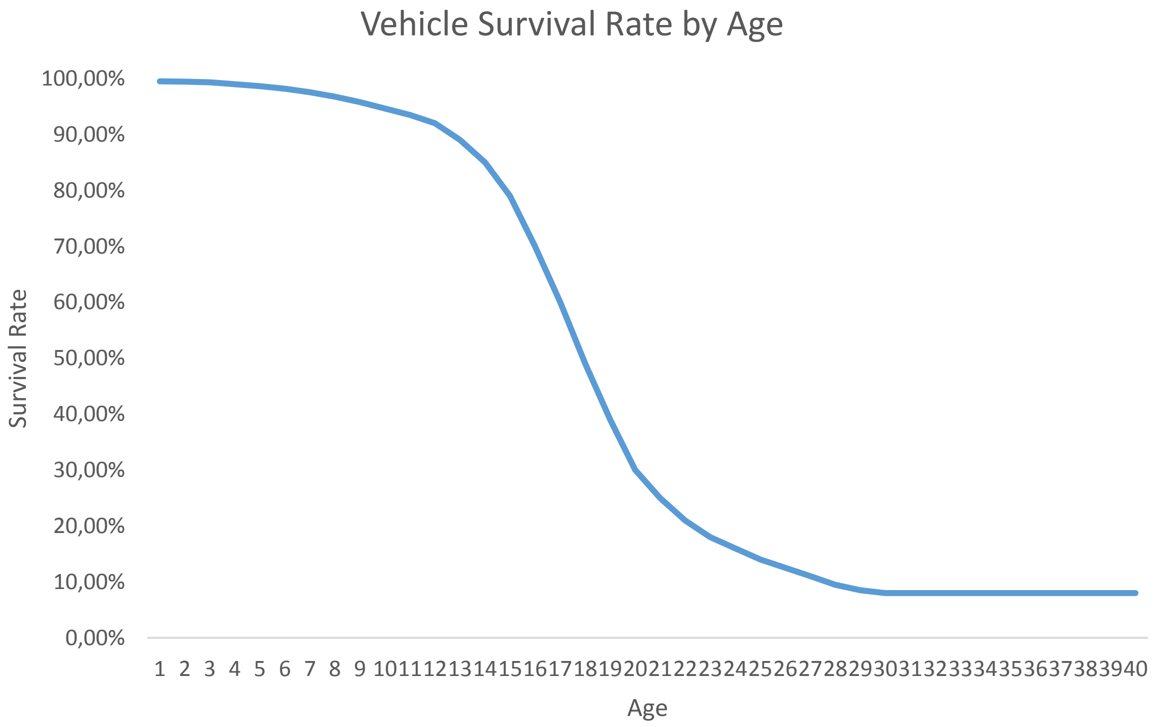
Appendix

Appendix 1 – Age vs. Mileage relationship

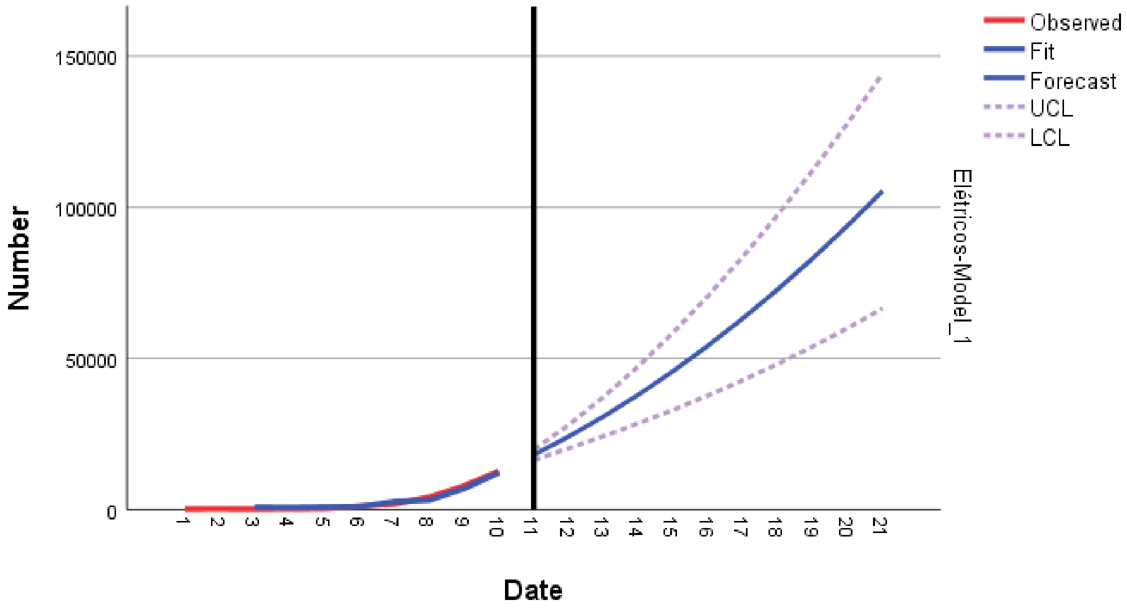
16. Número médio de quilómetros acumulados por idade dos automóveis:

Idade	Nº médio de quilómetros acumulados por automóveis
0 anos	7 471
1 ano	16 727
2 anos	32 274
3 anos	50 116
4 anos	65 914
5 anos	78 669
6 anos	91 690
7 anos	101 010
8 anos	107 089
9 anos	133 434
10 anos	135 145
10 - 15 anos	154 330
Mais de 15 anos	167 194

Appendix 2 – Vehicle Survival Rate by Age

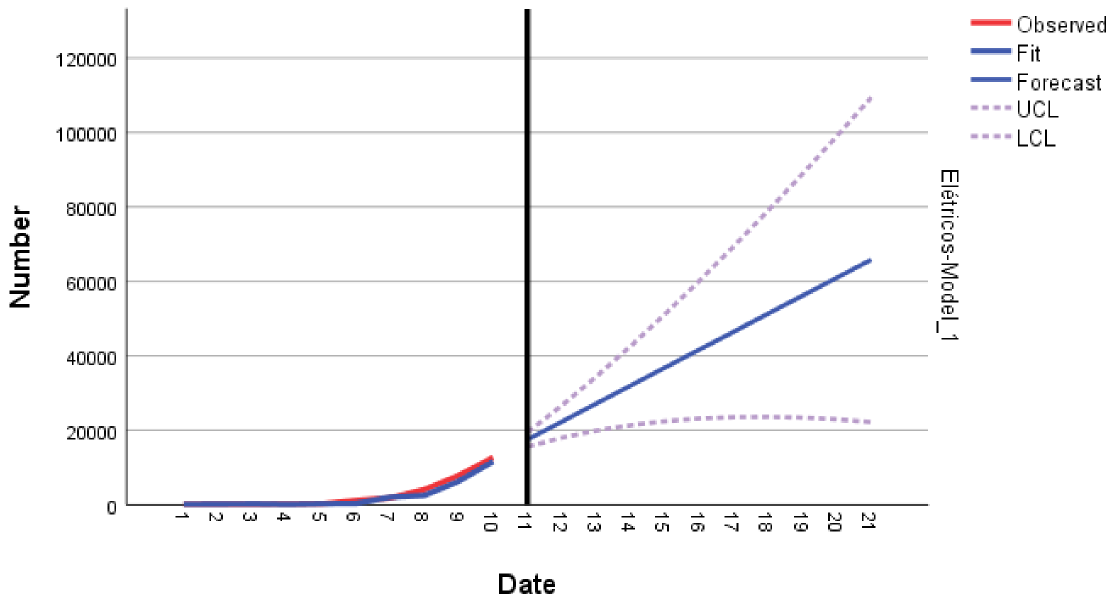


Appendix 3 – ARIMA model for Segment 9



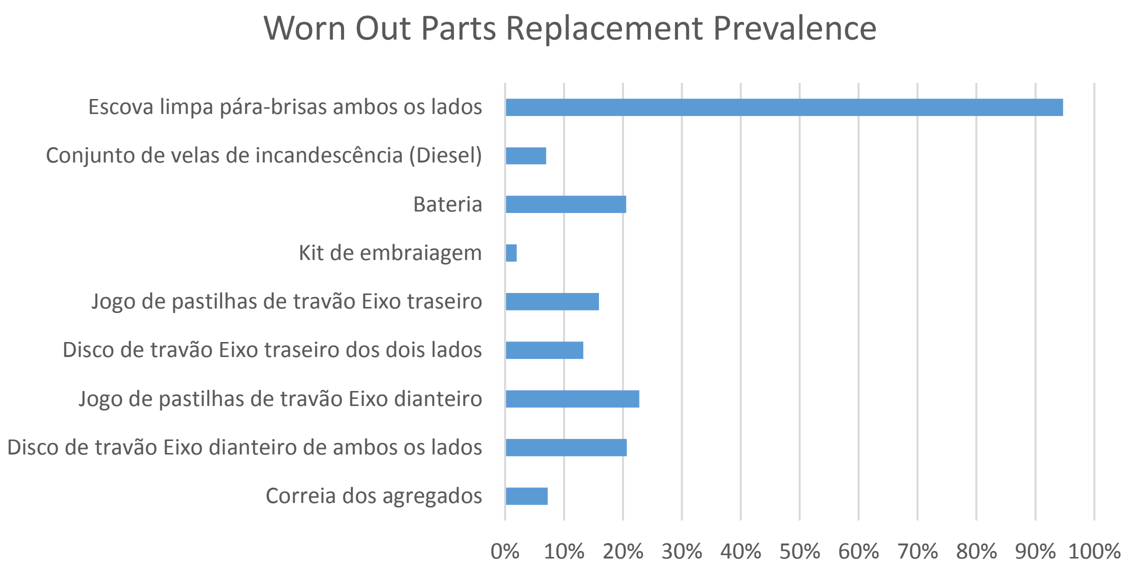
The plot includes observed cases, along with a fit line, forecast and upper and lower confidence levels. On the x-axis are plotted consecutive years since the beginning of the analysis, with 21 corresponding to the year 2030.

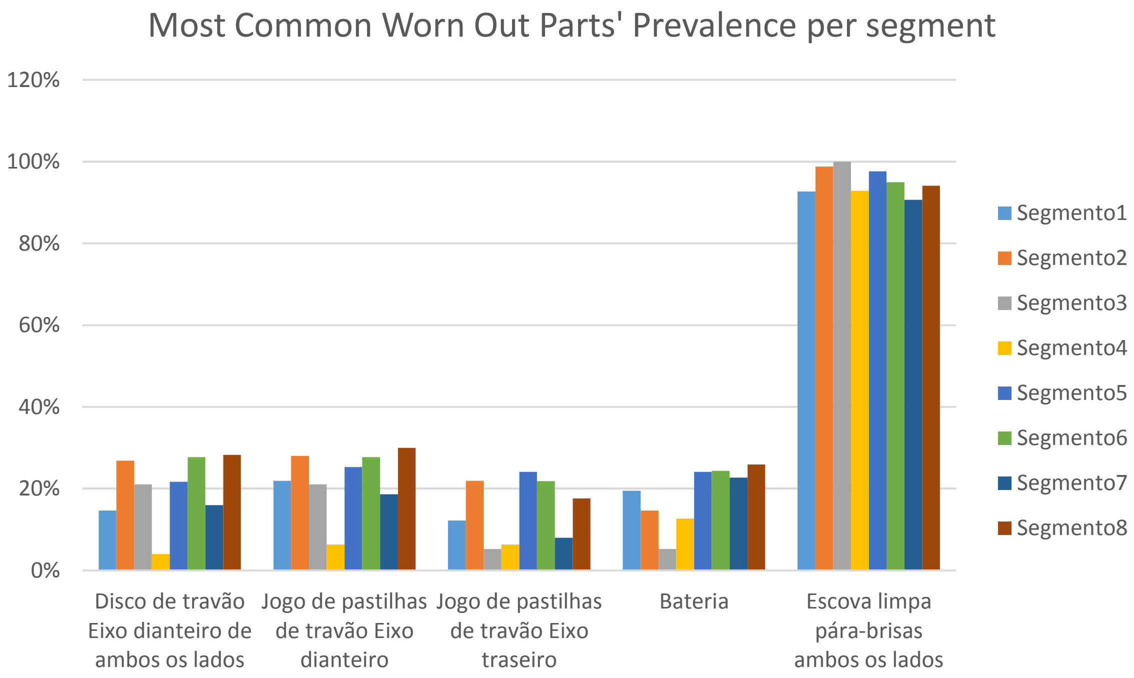
Appendix 4 – Brown's Linear Exponential Smoothing model for Segment 9



The plot includes observed cases, along with a fit line, forecast and upper and lower confidence levels. On the x-axis are plotted consecutive years since the beginning of the analysis, with 21 corresponding to the year 2030.

Appendix 5 – Details on Most Commonly Worn Out Parts





Appendix 6 – Details on Most Commonly Replaced Parts